

White Paper Report

Report ID: 98614

Application Number: PK5007208

Project Director: Sarah Kansa (skansa@alexandriaarchive.org)

Institution: Alexandria Archive Institute

Reporting Period: 1/1/2009-3/31/2011

Report Due: 6/30/2011

Date Submitted: 7/1/2011

Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment

White Paper to the NEH Division of Preservation and Access
Advancing Knowledge: The IMLS/NEH Digital Partnership
Grant #: PK-50072-08

1/1/2009 – 3/30/2011

June 27, 2011

Sarah Whitcher Kansa¹

Alexandria Archive Institute
skansa@alexandriaarchive.org

Eric C. Kansa²

Alexandria Archive Institute
& UC Berkeley, School of Information
ekansa@ischool.berkeley.edu

This work is licensed under the Creative Commons Attribution 3.0 Unported License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>



ABSTRACT

The Alexandria Archive Institute received a grant of \$250,609 from the National Endowment for the Humanities (NEH) and the Institute of Museum and Library Services (IMLS) for the project *Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment*. The project observed creators and users of online cultural heritage collections to determine how they use current tools in their research and to identify areas where innovations can improve Web-based scholarship. In collaboration with the Information and Service Design (ISD) Clinic at UC Berkeley's School of Information, we worked intensively with data creators and users in small working group settings to document their needs and experiences working with cultural heritage data sharing systems. We drew on the diverse experiences and insights shared by representatives from multi-organizational and interdisciplinary stakeholder groups, including museums, active field projects, public archaeology, cultural resource management, specialists, and junior and senior scholars, to explore how current Web technologies may better meet the needs of these groups. This research guided developments to Open Context (<http://opencontext.org>), an online data publication system for archaeology, small museums, and other field researchers. Other outcomes include guidance for “good practice” in developing Web-based collections and data publication systems. This report details activities and accomplishments over the course of the project, from January 2009 through March 2011.

¹ Project Manager

² Technology Consultant

1. INTRODUCTION

Understanding user needs and user experience requirements represents an essential foundation for the successful development and deployment of computing systems to enhance humanities research. Web-based technologies have the potential for making cultural heritage content freely and globally accessible, and easier to browse, find and analyze. However, such technologies will have limited impact for enhancing scholarly productivity and quality if they do not account for human and social factors in their application and use (Dempsey 2006, Raben 2007). Such factors include workflow patterns, incentives and reward structures, collaborative processes, and competitive concerns (Carr 2006).

This project helps complement some recent studies in user needs (especially Harley et al. 2010) by focusing more on specific implementation concerns with working systems and less on more general concerns in scholarly user needs. While recent studies and publications now provide a much better understanding of researcher user needs, the technical and social context of research continually evolves. This evolution means that user needs should be continually tracked and translated into practical and functioning systems.

2. PROJECT GOALS & ACTIVITIES

2.1 Goals: Problems and Solutions

Investigations undertaken in this study have resulted in guidance to help humanities computing efforts to best meet the diverse user needs in online data sharing systems. User experience optimization cannot easily be conducted in the abstract; it requires iterative cycles of development, deployment and evaluation with functioning systems. To this end, this project employed extant systems and tools, including: Open Context (<http://opencontext.org>), an open access data publication system for archaeology and field science disciplines; and open source collection management tools provided by Omeka (<http://omeka.org>). Prior to this project, we had already has validated key data models and technologies needed to support field and collections data discovery, browsing, and exploratory analysis using Open Context. By using working systems, this project was able to explore some specific design and work-flow concerns.

To help guide our efforts in improving these existing systems, this project looked at many prior efforts to deliver cultural heritage to the research community via the Web. This research was in part motivated by outcomes of a workshop organized by this project. Many workshop participants reported that Google, Google Scholar, and JSTOR were important resources, though some, especially in the public archaeology sector complained about lack of access to JSTOR (which requires an individual or institutional subscription). While these online tools and services now see heavy use by scholars, strikingly, participants uniformly said that they used few, if any Web-based collections or resources developed by the research community. The lack of uptake of these Web-based collections clearly needed study.

As discussed below, we came to recognize that issues that limited researcher uptake of existing Web-based collections had much to do with Web architecture concerns. Web architecture and technical systems design is not a traditional focus of “user needs” studies. However, our study of user needs identified clear demand for greater discoverability and interoperability of cultural heritage collections. Web architecture limitations of many existing systems hamper uptake and use because these sites are hard to find, and their content is hard to relate with content from other sources. Many sites seem to have aimed for appearance and front-end design as “destination websites,” but lack machine-readable representations of data or even structures that facilitate linking and indexing by search engines. Thus, over the course of this project, we became increasingly concerned with back-end technical architecture issues rather than front-end user interface issues.

The primary contribution and outcomes of this study therefore focus less on the front-end user experience, but rather on service design and Web architecture principles needed to power compelling user experiences. If systems have well designed Web architectures, they will be far more flexible in supporting different forms of front-end interfaces while better supporting discovery and aggregation. In this sense, our design philosophy emphasizes that user experience needs to be considered more broadly than how interactions work in a single site. Systems need to be open and flexible enough to be experienced through third party interfaces. Systems that can be experienced through alternative interfaces and developed by third parties that may index, aggregate, and “mashup” resources from a variety of sources can better support iterative experimentation in what works best to meet user needs. Thus front-end experience issues cannot be segregated from backend architecture concerns.

To promote better Web architecture practice in deploying Web-based resources for research, we developed a set of guidelines and functioning Web services / APIs that illustrate key design principles. The perspectives and methods for data dissemination and knowledge creation developed here have wide application across the multiple disciplinary boundaries that make up different strands of cultural heritage preservation and research. Indeed, methods and design patterns that successfully promote effective collaboration in working with structured data represent an important contribution with significance beyond the humanities.

2.2 Implementation

Implementation of this project involved drawing upon the experiences and insights of representatives from different humanities communities who regularly work with collections. This international group of representatives, hereafter called “Community Liaisons” represented museums, universities, cultural resource management and conservation organizations, public archaeology, and independent scholars. This project proposed an iterative process of assessing user experience. Over two years, the project cycled twice through a process of assessment, technology development, user testing, feedback and assessment. The Alexandria Archive Institute provided overall management and administrative coordination for the project. The Information and Service Design Clinic at UC Berkeley’s School of Information documented user needs and user experience requirements and developed specific implementations based on these requirements through enhancement of Open Context.

Central to evaluating user experience is removing the subjectivity of the developers of a system and focusing instead on the needs and diversity of the end-users. To achieve this, evaluations must draw from the experience and input of users with different backgrounds that share a common understanding of the site (Rubinoff 2004). Most current efforts at sharing digital heritage are taking a “top down” approach, drawing on the expertise of senior scholars to determine community needs. This project took a different approach, inviting scholars who are not yet senior to become active participants in creating the end-user experience. The Community Liaisons who participated in this project were, thus, our most valuable assets because their backgrounds, experiences and expectations have guided the outcomes of this project.

Our approach: In researching user needs, this project emphasized principles of participatory design. One of the best ways to understand user needs is through richly textured and highly contextualized examination of how users work toward their professional goals. Four major approaches were used in this investigation:

- *Plan from a Future Scenario:* We worked with the project Community Liaisons to learn how they envision an ideal social and technical framework for meeting their needs (their “dream tool”). Once an ideal system was described, we asked team members to identify specific developments and changes (in technology and practice) that need to be accomplished for that future to be realized.

- *Conduct Activity-Focused Research:* Many aspects of user experience are difficult to identify or articulate in the abstract. We observed and interviewed participants as they performed professional tasks so that we could understand their workflows.
- *Understand Context:* Online data sharing systems and collaborative tools like Open Context do not exist in a vacuum. There are several other systems such as email, electronic journal repositories, social software (blogs, tagging systems, social networking sites), and library resources that support research and collaboration. This project sought to understand how new humanities computing and collaboration systems fit into the larger ecology of online and offline exchanges. Noting the significance of the “ecology” of online services and resources motivated our emphasis on backend service design issues.
- *Platforms and Customization:* One of the most important goals of exploring user experience is to identify common or important patterns that will inform technical requirements. This project explored the underlying goals and needs of the project Community Liaisons in order to develop powerful but flexible solutions that can be cost-effectively adapted for different users. Again, adaptability and design flexibility must rest on a firm foundation of good Web architecture design.

2.3 Activities

- *Workshop:* This project launched with a 2-day workshop entitled “Exploring User Needs around Digital Heritage,” which took place January 23-24, 2009 at the School of Information at UC Berkeley. This workshop brought together the twelve Community Liaisons participating on the grant, as well as additional participants, to develop specific strategies and methods to enhance access to and usability of primary archaeological research content. Discussions focused on primary content, which we see as a promising but overlooked area of scholarly communications. The workshop also explored how different communities communicate and collaborate and how they are (or are not) using digital resources in their work. We discussed how better access to research might help address new questions in the different communities represented. The workshop concluded with participants forming small groups to identify some tangible next steps to improve access to and quality of primary content within and beyond their identified community. (See Appendix I for a workshop agenda and list of participants.)
- *Small Groups / Case Studies:* Project Community Liaisons formed small groups based on commonalities in what they identified as their “community” and their primary interests and concerns around data sharing. These small groups explored ways that they could move their communities toward greater facility in data access to improve research opportunities. Each group developed goals for the end of Year 2 (described in section 3.1.6). The four small groups were as follows (and see Appendix II for an overview of the goals and planned outcomes for each small group):
 - Data-heavy, institutionally-held collections, where content is maintained by an institution. The institution may lack internal technical resources to disseminate that content and tools are needed to help expand the reach and usability of data.
 - A blind analysis of “other people’s data” to demonstrate possible research benefits of data sharing by addressing new research questions with old data. Analysis of other people’s data also helped informed us about key metadata and data documentation needs required to make legacy data usable for research purposes.
 - Media-heavy, collaboratively developed collections. Some project participants sought to cultivate collaborative ties and information sharing in emerging areas of specialization. They wanted to develop semi-formal online collections for sharing identifications and type-specimen, spreading best practice, social networking, and sharing publications. These “bottom up” initiatives may have little or no institutional technical support or dedicated funding. They needed simple open source tools and services that they can manage and deploy on their own.

- Large excavation datasets, where data from multiple analysts over many years needs to be published, integrated and cited. Disseminating excavation datasets required managing highly (though idiosyncratically) structured data and sometimes large image and media collections. Participants on this project that shared large excavation datasets primarily wanted to use online data sharing tools to complement and enhance their narrative publication goals. Case studies informed developments to Open Context to improve data sharing workflows and data discovery.
- *Environmental survey of current heritage data sharing initiatives:* Over the course of the two-year project, we documented 68 websites to help us understand current uses of the Web for sharing cultural heritage content. Our goal was to determine whether some commonalities exist among sites that appear to be “successful” (such as reaching a wide audience and facilitating reuse of content by others). Each site was approached using the same set of criteria. The criteria and a complete list of sites reviewed are available in Appendix 3. Based on these reviews, we undertook a case study focusing on the development and sustainability of websites related to the cultural heritage of Iraq over the past 20 years, which will be published in 2011.

2.4 Changes to the Project

Our original proposal planned for two workshops over the course of the project. However, our first workshop in Year 1 resulted in the creation of four smaller groups collaborating on different research topics. Since, like all scholars, our participants face tremendous time and attention constraints, we chose to meet with the small groups individually in Year 2 so that we could maximize our time with them and focus on their specific topics. These small group meetings were highly productive and resulted in the various project accomplishments described below.

3. ACCOMPLISHMENTS

3.1 Outcomes

3.1.1 Rationale for Data Sharing as Publication

A primary concern that emerged from our discussions with researchers is a disconnect between incentives and workflow concerns for the dissemination and consumption of online research data. That is, consumers of online research data often have expectations about those data (particularly with regard to discovery and interoperability) that are difficult to meet given the motivation, time, and effort constraints of data producers. To better cultivate data sharing in archaeology (and beyond), the needs of data creators and the needs of data consumers need to be reconciled and balanced.

- *Data Dissemination Needs:* By virtue of their involvement in this study, participants generally favored data sharing. However, they identified several obstacles such as time, resource, and attention challenges. Many managed large datasets and felt that they never had enough time or attention to adequately fix problems and inconsistencies. They did not want to be over-burdened with metadata and data documentation. Some participants felt that narratives and interpretations should accompany data dissemination. Others wanted to gain more exposure for their work and organizations, and saw publishing Web-based collections as a good strategy. They felt their collections could be readily understood even without extensive narratives and interpretations. On the data dissemination side, researchers were much more concerned with the quality and integrity of their own data, and how these data can be visualized or expressed to support narrative interpretations. We note that data creators who participated in our study did not prioritize relating their own data with data from other sources.

- *Data Consumption Needs:* Workshop and small group discussions over the course of the grant indicated that researchers desire access to large amounts of high-quality data. They value standards, data documentation, provenance and reliability highly. Project participants wanted access to data for numeric / statistical comparison with their own data as well as resources for finding comparanda (especially drawings and photos). Access to online reference collections and discovery of comparanda were especially important for researchers working in emerging areas of specialization (such as micro-morphology and microarchaeology). Data integration (pooled analysis of data from different sources) was frequently raised as a feature of a participant's "dream tool." However, data quality needs varied; some wanted high quality data to analyze, while others showed greater willingness to put up with poor data quality so long as they had access.

In exploring the above needs, we began to understand that the requirements for efficient data discovery and use sometimes clashed with concerns over time and resource burdens expressed by data creators. Efficient data discovery and reuse typically requires effort to document data and to align data to common standards, but data creators expressed some resistance to expending such effort.

The researchers that participated in our study had great familiarity with certain desktop data management applications, particularly Microsoft Excel and, to a lesser extent, relational databases (Access and FileMaker). These desktop applications proved to be almost universally used by our project participants. Use of Web-based data management and collaborative tools such as Google Docs was less frequent. Understanding of Web technologies and standards remained well outside of their realm of expertise.

Our work with project participants suggests little motivation or interest in having researchers "markup" their own data to align these data with more general Web or semantic standards. Project participants generally saw this as a somewhat abstract goal, disconnected from their immediate needs, and usually felt such semantic and standards alignment stood too far outside of their area of expertise. For example, we had only limited success in working with or researcher team to use Open Context's schema mapping tool. This schema mapping tool ("Penelope") guides users in aligning their source data to the data standard used in Open Context (ArchaeoML). The tool is rather complicated and needed continued developed to accommodate highly heterogeneous source data. Use of Penelope required telephone calls to walk users through its interface and guide schema mapping. Because Open Context's editors did not feel comfortable with users uploading data by themselves without checking on the quality and soundness of schema mapping, we decided not to continue with user experience refinements with Penelope. Instead we repositioned Penelope to be a tool used by Open Context's editorial staff as part of an editorial workflow. User interface refinements focused on the public site of Open Context.

At this time, we believe that it is very difficult for researchers themselves to relate their data to sophisticated schemas/ontologies (such as ArchaeoML). However, such standards are needed to make data easier to discover and use on the Web. Because of these conflicting issues, we believe that *editorial intervention* may best produce the quantity and quality of needed metadata without overtaxing individual data contributors.

3.1.2 Guidelines for Sharing Digital Collections

This project has produced two sets of guidelines to help scholars who wish to share digital content and those who wish to use digital content navigate the current landscape. These documents, though drawing on examples from archaeology and describing data preparation for publication in Open Context, can be easily generalized to inform the broader digital humanities about simple ways to prepare (from the author's perspective) and evaluate (from the user's perspective) datasets distributed on the Web.

- *Guidelines for Web-Based Data Publication:* The environmental survey described at the end of section 2.3 highlighted the lack of awareness of technical and architectural aspects in Web design within much of the digital humanities. Only one quarter of the reviewed sites had content in a format that made it “easy to reuse” (such as clear licensing information, open formats, stable URIs, and machine-readable data). In response, we developed a document, which we distributed at conferences and on the Web that provides guidelines for humanists who seek to share digital content on the Web. The document provides simple tips and examples of appropriate uses of the Web for sharing content. The document is meant to guide projects that are building websites or are revamping extant sites. The aim of the document is to raise awareness of good practices for sharing machine-readable content. The document highlights why good machine interfaces are just as important as human interfaces for the long-term usability and longevity of cultural heritage content. The document outlines several simple standards and design approaches researchers and collections managers can use to ensure that their websites follow good design principles from the outset. In addition to design tips, the document includes advice on how to approach digital resources to determine their quality and usefulness. This document is available at <http://ux.opencontext.org/blog/good-practice>.
- *Data Publication Guidelines for Contributors to Open Context:* In response to the National Science Foundation’s (NSF) recent data management plan requirements, and based on research conducted during this project, we developed specific guidance on how scholars can prepare datasets for publication with Open Context and how to budget for data dissemination and archiving (<http://opencontext.org/about/publishing>). Open Context also points to the guides to good practice³ prepared by the Archaeology Data Service and Digital Antiquity. These publication guidelines are meant to help inform and improve researcher data management and documentation practices from the very inception of a project. Improved data quality (with normalization and validation to reduce errors and improve consistency) will make it easier for data publishers like Open Context or Digital Antiquity to offer usable, machine-readable data and to link data to community semantic standards. By improving data management at the outset, we believe some of the workflow and incentive barriers toward making data public can be reduced.

3.1.3 Technical Developments

Based on discussions with researchers and observing the project participants as they grappled with various ways of sharing content, we enhanced Open Context in a number of ways. Many of these enhancements focused on back-end Web services to promote aggregation and reuse of data. Deploying these back-end services provides functioning examples of machine-readable services to illustrate issues discussed in the Guidelines documents:

- *Faceted Search:* The “hit or miss” nature of Boolean keyword searches limits the effectiveness of information repositories for meeting scholarly needs, especially with regard to meta-analyses attempting to look across different datasets. In faceted navigation applications, users explore multiple dimensions of hierarchically-structured metadata with simple and intuitive “point and click” selections to progressively home in on specific information from a larger collection. Because filters are applied across an entire collection, users have greater certainty in the comprehensiveness of their results than with keyword searches. Feedback, in the form of subtotals for the numbers of items falling under each available facet, guides users in selecting additional filters. This feature offers users cues about the size and composition of the collection they are searching (Hearst 2006).
- *Technical Interoperability:* Researchers in our study wanted access to large amounts of data, ideally with comprehensive coverage. Given the effort involved in compiling large quantities of high-quality data, we believe that effort needs to be distributed. Interoperability between data providers needs to exist in order to support researcher goals for comprehensive coverage. Many of Open Context’s enhancements during this grant period focused on interoperability through deployment of RESTful (“Web style”) Web services

³ <http://ads.ahds.ac.uk/project/goodguides/g2gp.html>

(Atom feeds, JSON data, and KML). In addition, we are moving toward greater support of “Linked Data” methods of interoperability by publishing data using RDFa and referencing outside vocabularies through URIs. See <http://opencontext.org/about/services> for documentation of these services. These services provide working examples of machine-readable data interfaces advocated in our “good practice” guides.

- *Data Citation Issues*: Researchers clearly demanded permanence and stability in online data. Data curation and archiving are crucial needs. To respond to these demands we are working with the California Digital Library to use their persistent identity (and other data curation) services and implement the DataCite standard (see section 3.1.4).
- *Searchable Documentation*: Review of server logs and discussion with project participants showed the importance of documentation about Open Context, especially its editorial policies, goals and services relating to the new NSF data management plan (see section 3.1.5). We added enhanced search functionality to Open Context, including “highlighting” of search terms in text snippets to better meet these needs.
- *Linked (Open) Data and Shared Vocabularies*: Participants wanted freedom to define their own semantics for their own projects, but also expressed interest in using more generalized, community standards for using other people’s data. This represents a tremendous challenge, and Open Context development hopes to close this gap by enhancing publishing processes to reference standard vocabularies (where applicable) while retaining a dataset’s original descriptive terminologies. An example dataset referencing standard vocabularies can be found [here](http://pleiades.stoa.org) (standard vocabularies are from the NEH-funded Pleiades Gazetteer, <http://pleiades.stoa.org>). This example illustrates a set of coins from a site in Turkey where Open Context editors related mint-marks on the coins with place-entities (referenced by URIs) from Pleiades. These assertions are expressed in RDFa⁴ using an annotation method advocated by Pelagios (<http://pelagios-project.blogspot.com/>), a JISC-funded collaboration of several digital humanities projects, including Open Context. This example demonstrates a “Linked Data” approach, where multiple Web-based resources can be linked together to enhance data discoverability and data analyses through URI reference to shared concepts.

3.1.4 California Digital Library Archival Support

As discussed above, researchers want electronic resources to have greater permanence, including reliable and persistent accessibility. Data preservation and ensuring the long-term persistence of citations represent tremendous institutional challenges. These challenges are well beyond the capacity of a small effort like Open Context. However, as in many areas, emerging distributed services and infrastructure can help data sharing efforts like Open Context meet such challenges. Open Context has begun to integrate data preservation and curation services from the University of California’s California Digital Library (CDL) as part of the library’s participation in the NSF’s DataNet initiative. These include:

- *Minting and binding of ARKs (“Archival Resource Keys”)*: ARKs are special identifiers managed by an institutional repository. The CDL will help ensure the objects associated with these identifiers can be retrieved in the future, even if access protocols such as HTTP change. Open Context now binds its Web identifiers to CDL-minted ARKs to facilitate long term retrieval and citation of content.
- *Data archiving*: The CDL also provides data curation and stewardship to maintain integrity of digital data and to migrate data into new computing environments as required.

We should note that we are currently working with developers of the CDL’s Merritt service (<http://www.cdlib.org/services/uc3/merritt/>) to accession content from Open Context to the CDL’s data preservation infrastructure using Atom feeds. Open Context publishes Atom feeds of its content, and Merritt has

⁴ We gratefully thank Sebastian Heath (New York University / American Numismatic Society) for his guidance in expressing such annotations in RDFa.

successfully read and “crawled” these feeds to retrieve content for curation. Thus, the kinds of Web architecture approaches advocated by this project not only facilitate collections discovery, aggregation and reuse, but also digital preservation.

3.1.5 NSF Recognition

In 2010, the NSF announced new data management requirements for grantees. Grant-seekers are now required to supply data management plans as supplementary materials to their proposals in order for their proposals to be accepted. This new requirement has the potential for improving transparency in research and facilitating new research programs that bring together results from multiple projects. Many grant seekers will probably lack expertise and technical support in making data accessible. Open Context offers grant-seekers fee-for-service data publishing and archiving services (through the CDL), as well as guidance in preparing a data management plan. In October 2010, the NSF listed Open Context as an example of a data repository to help guide grant-seekers in the archaeology program (see example here: <http://www.nsf.gov/sbe/bcs/arch/senior.jsp>). NSF’s data management policy took effect on January 18, 2011, and the first archaeology proposals under this new regime are due in mid-summer 2011. To date, a dozen proposals have requested estimates from Open Context for such services.

3.1.6 Small Working Group Outcomes

A variety of Web-based outcomes have resulted from the small group projects that we identified based on the workshop at the inception of this project. These results are primarily aimed at demonstrating the impact of digital content that is shared using tools and design principles that maximize the potential of the Web. Descriptions of the small groups and their goals are provided in Appendix 2. The results of the small group projects are as follows:

- *Content-Sharing and Community-Building for Microarchaeology:* Community Liaison Catherine Foster developed MicroCommons (<http://microcommons.org>), an Omeka-based community portal for sharing and discussing research on archaeological microdebris. MicroCommons, the first of its kind in the nascent field of microarchaeology, holds great promise for advancing research in this area and exposing microdebris scholarship to related fields. Foster built MicroCommons using the open source collection management tools offered by Omeka. She felt that Omeka offered an excellent base for her project because she had a sufficient level of technology experience to feel confident building upon Omeka’s tools to create a more custom-built site to meet her needs. Her site emphasizes informal sharing and an attempt at collaborative development of online reference collections in an emerging field.
- *Digital Content from Iraq:* Community Liaison Eleanor Wilkinson developed a site to document and share archaeological and museum content from the region around Nineveh in northern Iraq. She trained an Iraqi archaeologist based at the museum in Mosul to work with the site and upload content. Because she did not have access to technology expertise to build upon Omeka (as in the MicroCommons example above), Wilkinson chose to model her site on another Omeka-based project developed with NEH support by the Alexandria Archive Institute in collaboration with the UC Berkeley School of Information, the Modern Art Iraq Archive (<http://artiarq.org/maia>). By using the same template and tools developed for the MAIA project, Wilkinson was able to achieve the custom features she needed without having to put additional funds into hiring a developer.
- *The Badè Museum and the San Diego Archaeological Center:* Project participants from the Badè Museum and the San Diego Archaeological Center (SDAC) sought to increase awareness and use of their collections by providing detailed, open Web access to them. They also sought to maintain institutional branding of their online collections. In order to meet these needs, we established a workflow for them to publish their collections to Open Context and, via JSON-expressed data, channel collections back to their institutional websites. Using this simple process, scholars managing institutional collections increase the exposure and

longevity of their content by Web-publishing it in Open Context, digitally archiving it (with the CDL), and maintaining an updated database of their content on their institutional website via feeds from Open Context.

- *Presidio Archaeology Laboratory*: The Presidio Archaeology Laboratory sought to explore options for increasing public engagement with their collections while maintaining correct associations among the multiple facets of a project. We worked with them on a pilot project, Building 39 (see <http://opencontext.org/projects/C5B4F73B-5EF8-4099-590E-B0275EDBA2A7>), which produced multiple analyses and datasets. This served as a proof of concept for the Presidio Archaeology Lab to feed their digitized collections back into their site (like the Badè Museum and the SDAC examples described above).
- *BoneCommons*: BoneCommons (<http://bonecommons.org>) is a discipline-specific website built with Omeka that serves the international zooarchaeological community. Much like MicroCommons, BoneCommons offers a space for sharing ideas, publications, questions, and comments around a sub-disciplinary topic. Over the course of this project, we worked to develop new features for BoneCommons to draw in content of interest to zooarchaeologists from across the Web. Currently, BoneCommons features a “Datasets” page that draws feeds from projects in Open Context that have “zooarchaeology” as a keyword. Users can search within those zooarchaeology-related datasets and retrieve results without having to navigate away from the BoneCommons site. This Datasets page can draw feeds from any Web resource that provides feeds of zooarchaeological data so that multiple Web resources can be viewed and searched together. Unfortunately, few of these exist. Thus, a portion of the grant period was dedicated to informing the zooarchaeological community about the potential of Web-based data sharing and ways to share to maximize the reach and reuse of content (such as via Open Context’s API); and to encouraging zooarchaeologists worldwide to share their research on the Web.⁵
- *A Model for Data Publication*: To help explore where documentation effort can be most effectively spent, one of this project’s small working groups considered data sharing from the perspective of “end users” who consume legacy data for reuse and analysis. We worked iteratively with multiple archaeological projects that included multiple specialists contributing different forms of analytic data. Members of these large, multi-specialists teams, especially the project directors, tended to regard data dissemination as an important aspect of their larger publication goals. For publication goals, data citation represented the key user need. Researchers working with larger datasets needed support in associating analytic data with media resources (images, scanned drawings, and field notes). Some researchers felt the effort required for cleaning up errors and inconsistencies in data made data sharing too costly. Software and help to improve data quality prior to public dissemination may help mitigate this concern.
- *Preliminary Exploration of Data Re-Use*: Three project participants with expertise in zooarchaeology conducted a blind analysis and report-writing study, using a publicly available faunal dataset from Chogha Mish, Iran. This study demonstrated that even sparsely documented datasets can show sufficient integrity to merit further analysis. Thus, though detailed documentation of digital collections is ideal, collections with little documentation should not be discounted. In exploring how researchers tried to understand and analyze “other people’s data,” we hoped to inform better data publishing practices.

3.2 Lessons and Recommendations

Over the course of this study, we have identified two critical issues for creators and users of Web-based digital humanities content. Below, we describe these issues and offer recommendations for how they can be ameliorated.

⁵ Advocating for data sharing on the Web and adherence to approaches that promote integration and informed reuse took place mainly in the context of the 2010 International Council for Archaeozoology (ICAZ) conference and subsequent publications on these topics. We also campaigned for increased sharing of poster and oral presentations at the ICAZ conference and, to date, over 120 scholars have shared their work on BoneCommons (see here: <http://alexandriaarchive.org/bonecommons/items/browse/tag/presentation+shared>).

Issue 1: Many Needs, Many Contexts: Some researchers (e.g. collections managers at the Badè Museum and the San Diego Archaeology Center) sought low-cost solutions for sharing their collections in order to improve their institution's online visibility. Some groups primarily needed to develop and access comparative collections. Others wanted to publish datasets to accompany and augment narrative publications, while still others wanted access to large datasets for comparative analyses. Finally, another group wanted to share collections to build community in emerging subfields of archaeology.

Recommendation: It is difficult to envision one single portal or system that could meet all needs. The Badè Museum and San Diego Archaeology Center participants wanted to share collections data, but also enhance their online Web presence. Web services delivering data from Open Context to their institutional websites helped meet these goals. Similarly, other types of software (blogs, content management systems, online journal systems) support narrative publication workflows. "Loose-coupling" of data management and data publishing systems like Open Context and other content management systems specializing in more narrative/ textual content can help meet some needs. We demonstrated this through integration with the VisiplePast.net (WordPress-based) system developed by Sorin Matei at Purdue University (Matei et al in press).

Large-scale aggregation and study of data published from a number of sources offers exciting research opportunities, but will be difficult to realize without access to "machine-readable" data from Web-based collections. While this assessment is by no means new, we recommend that the community start with simple measures for data portability rather than waiting for all the difficult issues of semantic alignment to be solved. By providing machine-readable data using relatively simple, open formats like JSON or Atom and other XML or RDF formats, data providers can encourage greater experimentation with merging different Web services and applications (including new visualization options) and integrating disparate collections. That experience can then inform semantic standards alignment. In addition, even if data publishers find it too difficult to provide Linked Data initially, if they adhere to some basic principles of Web architecture (especially stable URIs), they can at least make their data more "linkable" by others.

Issue 2: Aggregation and Discovery are Important, but are Poorly Supported: Many researchers participating in our study wanted to use comprehensive data resources collected by their colleagues for numeric analyses. Others simply wanted better aids to find comparative material. A recurring theme among project participants when discussing "dream tools" was easy, simultaneous access to many types of high-quality content (reports, datasets, images, maps). Unfortunately, many digital collections are locked in silos with very poor interoperability and Web architecture, making data discovery and aggregation difficult.

Recommendation: Fast and simple access to multiple types of content from multiple sources/projects will only be possible if all projects expose their content to the Web, especially through machine-readable formats. NEH and other granting foundations need to mandate at least minimal good practices and Web architecture for new digital collections. New collections need to better support linking (ideally to Linked Data), use valid open standards, and offer some simple machine-readable data such as Atom feeds and JSON. More sophisticated Linked Data technologies such as RDF, SPARQL, RDF-OWL, etc., can also be encouraged, though they may present some adoption barriers. Ideally, data providers should offer data in a variety of open formats, since Atom, JSON, RDF, (and now "Microdata"⁶) all have advantages for certain applications. As more researchers and developers engage with machine-readable data, expectations and needs for certain formats of data will no doubt evolve, and may even converge on certain standards.

⁶ Microdata is a new open format advocated by Google, Microsoft and Yahoo for expressing "machine-readable" data in Web resources. Schema.org has published several microdata vocabularies in different areas.

While machine-readable data is critical for reuse, it almost goes without saying that legal interoperability is also a critical concern. Our project participants saw great promise in analyzing data aggregated from a number of sources. However, they would need access and legal permissions in order to do so, even if data providers offered machine-readable data. Enabling large-scale research across multiple collections therefore requires open licensing (especially with Creative Commons) and access, at least of structured, analytic data.

While this project illustrates the need for certain key Web architecture approaches, we mainly focused on baseline accessibility of machine-readable representations. However, machine-readable data may be a necessary, but not a sufficient, requirement for aggregation and reuse. Likely, there may be areas where use of shared vocabularies (especially in Linked Data scenarios) will be more critical. To explore these issues and to help create incentives for opening Web based collections to more efficient third-party aggregation and reuse, NEH or other funders may want to consider programs specifically targeted toward *using other people's data*. Such programs can create practical experience in advancing the Web architectures and semantics needed to enable large-scale yet meaningful aggregation and study of cultural heritage collections. The NSF funded DataOne project may offer a compelling model for realizing these goals. The program may even specifically aim to provide resources to retrofit existing significant collections and improve their capabilities to offer machine-readable data and / or Linked Data.

Another critical need is to develop Web-technology skill sets within the researcher community (especially, in our case, archaeology). Archaeology has a small but growing number of data providers offering machine-readable services (see Appendix 3.3). To truly use these services to the maximum extent, the archaeological community needs more members with Web scripting and technology skills.

4. AUDIENCE, DISSEMINATION AND EVALUATION

4.1 Audience

The publications and presentations listed in section 4.2 illustrate the diversity of audiences this project has addressed. These audiences include digital humanists, museum professionals, archaeological researchers, scholars conducting field work, professional archaeologists working in cultural resource management, heritage managers working in public and government contexts, and scholars and students working in archaeological specializations.

4.2 Dissemination

4.2.1 Publications

The outcomes of this project have been disseminated in a variety of ways. In addition to the website reviews posted on the project blog the project has resulted in publications and conference presentations and workshops.

Print publications:

Kansa, Eric C. (in press) New Directions for the Digital Past. In *Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. (Accepted for publication: July 2010).

Kansa, Eric C. and Sarah Whitcher Kansa (in press) Towards a Do-it-yourself Cyberinfrastructure: Open Data, Incentives, and Reducing Costs and Complexities of Data Sharing. In *Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. (Accepted for publication: July 2010).

- Kansa, Sarah Whitcher and Francis Deblauwe (in press) User-Generated Content in Zooarchaeology: Exploring the "Middle Space" of Scholarly Communication. In *Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication*, edited by E.C. Kansa, S.W. Kansa and E. Watrall. Cotsen Institute of Archaeology Press: Los Angeles, CA. (Accepted for publication: July 2010).
- *Atici, Levent, J. Lev-Tov, Sarah Whitcher Kansa and Eric C. Kansa (in review) Other People's Data: A Demonstration of the Imperative of Publishing Primary Data. (Submitted to *Journal of Archaeological Science*, May 2011)
- Matei, S. A., Kansa, E C., and N. K. Rauh (In Press) The Visible Past / Open Context loosely coupled model for digital humanities ubiquitous collaboration and publishing. *Spaces and Flows*.
- Kansa, Sarah Whitcher and Eric C. Kansa (2011) Beyond BoneCommons: Recent Developments in Zooarchaeological Data Sharing. *The SAA Archaeological Record* 11(1): 26-29.
- McKechnie, Iain and Sarah Whitcher Kansa (2011) Transformations in Digital Communication and Collaboration: Recent Perspectives from Zooarchaeology. *The SAA Archaeological Record* 11(1): 10,29.
- * Buckley, Mike, S. Whitcher Kansa, S. Howard and S. Campbell (2010) Distinguishing archaeological sheep from goat bones using a single collagen biomarker. *Journal of Archaeological Science* 37(1): 13-20.
- Kansa, Eric C. and Sarah Whitcher Kansa (2010) "Mashable" heritage: formats, licenses and the allure of openness. *Heritage in the Digital Era*. London: Multi-Science Publishers, pp 105-112.
- Kansa, Eric C. and Ahrash Bissell (2010) Web Syndication Approaches for Sharing Primary Data in "Small Science" Domains. *Data Science Journal* 9: 42-53
- Kansa, Eric C., Tom Elliott, Sebastian Heath, and Sean Gillies (2010) Atom Feeds and Incremental Semantic Annotation of Archaeological Collections. In *Computer Applications and Quantitative Methods in Archeology – CAA 2010*, edited by J. Melero & P. Cano.
- Kansa, Eric C. (2010) Open Context in Context: Cyberinfrastructure and Distributed Approaches to Publish and Preserve Archaeological Data. *The SAA Archaeological Record* 10(5):12-16.
- Kansa, Eric C., Sarah Whitcher Kansa, Margie M. Burton, and Cindy Stankowski. (2010) Googling the Grey: Open Data, Web Services, and Semantics. *Archaeologies. Journal of the World Archaeological Congress* 6(2):301-326.
- Kansa, Eric C. and Ahrash Bissell (2010) Web Syndication Approaches for Sharing Primary Data in "Small Science" Domains. *Data Science Journal* 9:42-53
- Kansa, Eric C. (2009) Indigenous Heritage and the Digital Commons. In *Traditional Knowledge, Traditional Cultural Expressions and Intellectual Property Law in the Asia Pacific Region*, edited by Christoph Antons. Kluwer Law International: Alphen aan den Rijn, pp.219-244.
- Kansa, Eric C. (2009) Opening Archaeology to Mash-ups: Field Data and an Incremental Approach to Semantics. In *Digital Heritage in the New Knowledge Environment: Shared spaces & open paths to cultural content* (Conference Proceedings), edited by Metaxia Tsipopoulou. Hellenic Ministry of Culture: Athens, Greece.

* Published work contains link to online dataset(s) published in Open Context.

Online-only publications:

- Kansa, Eric C. and Sarah Whitcher Kansa (2010) Publishing Data in Open Context: Methods and Perspectives. *CSA Newsletter*. Vol. XXIII, No. 2 (September 2010). Bryn Mawr: Center for the Study of Architecture.
- Kansa, Sarah Whitcher and Eric C. Kansa (2009) Yes, it is all about you: User needs, archaeology and digital data. *CSA Newsletter*. Vol. XXII, No. 1 (April 2009). Bryn Mawr: Center for the Study of Architecture.
- Kansa, Eric C. and Sarah Whitcher (2009) Open Context: Developing Common Solutions for Data Sharing. *CSA Newsletter*. Vol. XXI, No. 3 (January 2009). Bryn Mawr: Center for the Study of Architecture.

4.2.2 Presentations

- 2011 Society for American Archaeology (Sacramento, CA)
 - *Planning for Archaeological Digital Data Management: Addressing the New NSF Requirement* (Forum co-organized with colleagues from [Digital Antiquity](#) and the [Archaeology Data Service](#))
 - *Effective Use of the Web to Support Archaeological Research* (Oral presentation)
 - *Lessons in Data Reuse: A Blind Analysis of Faunal Data from Iran* (Oral presentation)
- 2010 American Schools of Oriental Research (Atlanta, GA) *Publishing Archaeological Data from the Field to the Web* (Organized workshop)
- 2010 International Council for Zooarchaeology (Paris, France)
 - *Archaeozoology in a Digital World: New Approaches to Communication and Collaboration* (Organized session)
 - *Other People's Data: Blind Analysis and Report Writing as a Demonstration of the Imperative of Data Publication* (Poster presentation)
 - *Beyond BoneCommons: Recent Developments in Zooarchaeological Data Sharing* (Oral presentation)
- 2010 Digital Humanities Conference (London, UK) *The Modern Art Iraq Archive (MAIA): Web tools for Documenting, Sharing and Enriching Iraqi Artistic Expressions* (Oral presentation)
- 2010 Society for California Archaeology (Riverside, CA) *Exploring User Needs in Archaeological Data-Sharing: Case Studies from California Archaeology* (Oral presentation)
- 2010 Society for American Archaeology (St. Louis, MO)
 - *Of Glass Houses and Ground Stone: Open Data and Ground Stone Analyses* (Oral presentation)
 - *Practical Methods of Data Production, Dissemination, and Preservation* (Organized session)
 - *Quantity has a Quality all its Own: Archaeological Practice and the Role of Aggregation in Data Sharing* (Oral presentation)
- 2010 UC Berkeley School of Information (Berkeley, CA) *Carrots, Sticks, and Web Publishing of Scientific Data: Open Context in Context* (Seminar)
- 2009 Workshop: *Exploring User Needs around Digital Heritage*, Berkeley, CA (co-organized by the AAI and UC Berkeley School of Information)

4.2.3 Software and other Grant Products

This project released all software, documentation, reports, and content online and free-of-charge:

- *Software*: Open Context is an open source project licensed under the GNU-General Public License (version 3). The software is available for free download in the Google Code repository: <http://code.google.com/p/opencontext/>. Its component software (PHP, MySQL, Apache Solr) is released under related open source licenses.
- *Documentation*: Summaries of workshops and notes involved in the collaborative design process are publicly available under a Creative Commons Attribution License at the project's documentation site, [Heritage Bytes](#).
- *Cultural Heritage Content*: Field research and museum collections datasets related to this project are available through Open Context under a Creative Commons Attribution license. Because data reuse is a primary goal, the "no derivatives" license variant is not permitted. The noncommercial restriction is permitted in specific cases, especially in contexts where indigenous communities may regard such restrictions as required, but its use is discouraged in other contexts.
- *Guidelines*: Guidelines produced by this project are available for free online under a Creative Commons By license (<http://ux.opencontext.org/blog/good-practice>, <http://opencontext.org/about/publishing>). To

increase impact, we are currently preparing manuscripts for submission to open access, peer-reviewed e-journals such as First Monday or the Digital Humanities Quarterly.

- *Publications*: All publications associated with this project (listed in section 4.2.1) have been made available open access via self-archiving or publication with open access journals.

4.3 Evaluation

Recording metrics on these outcomes is challenging. Most of our feedback has been qualitative, based on an iterative process of technology developments and feedback from users. We have observed that online data is perceived differently than more narrative online content, such as PDFs of publications. The “primary” or “raw” nature of datasets causes users to hesitate in both posting and using them. How do they know a dataset can be trusted? How do they cite it? These issues impact all audiences mentioned above, regardless of the type of data being shared or accessed.

For this reason, we have chosen for continuing research to focus on data quality issues and formalized data publication in order to help datasets become easily accessed and used in the same way that print publications have become. As data publication becomes a more expected outcome of research, processes for formalized data sharing need to be in place. For instance, since the NSF began requiring data management plans of grant-seekers in January 2011 and placed a link to Open Context in their guidelines, visits to Open Context have increased. With escalating requirements and expectations, scholars will increasingly look for data dissemination venues and guidance. Thus, in addition to experimenting with building content-sharing tools for specific communities, our project outcomes aim to provide general guidance for how to better share data on the Web and examples of projects that have done so successfully.

To help document the impact of this project and its outcomes, we provide the following qualitative and quantitative measures:

- Open Context Use and Evaluation: Sebastian Heath (New York University / American Numismatic Society) recently discussed the challenges in reusing data published in Open Context at the Society for Scholarly Publishing (SSP) conference in Boston (2011). While he praised the data access provided by Open Context, he noted data quality problems. This feedback further supports our move toward “data publishing” models with editorial supervision to improve data quality.
- AnthroDPA: Eric Kansa participated in the NSF, NEH, and Wenner-Gren sponsored “AnthroDPA” workshop. Open Context is referenced as a data dissemination tool, and Eric authored a report regarding intellectual property issues in anthropological data dissemination. Participation in this workshop helps demonstrate Open Context’s significance and impact in informing data sharing and archiving efforts within anthropology and archaeology.
- [“Rome wasn’t Digitized in a Day”: Building a Cyberinfrastructure for Digital Classicists](#) (by Alison Babeu, Perseus Digital Library) discusses Open Context and its role and potential role for Classical archaeology.
- ThatCamp Mashup: Open Context was listed as a data source offering machine-readable data for a “mashup competition” in the London ThatCamp (2010).
- International recognition, through invited speaking engagements to: Hellenic Ministry of Culture, Greece (2008); American School of Classical Studies, Greece (2009); German Archaeological Institute and Cologne University / Arachne Project (2011); Institute for the Study of the Ancient World / NYU (2009)
- 30 different scholarly sources now reference Open Context (according to Google scholar)
- National Science Foundation now (as of Sept. 2010) lists Open Context as an example of a venue for grant-seekers to share their data as part of the new required data management plan.

- National Endowment for the Humanities (as of June 2011) lists Open Context as an example of a venue for data storage to fulfill the new NEH data management plan requirement (see here: <http://www.neh.gov/ODH/LinkClick.aspx?fileticket=jQ44xoe2ZjU%3d&tabid=108>).
- Linking: We hesitate to provide this metric, since it represents a crude measure of impact. According to Yahoo, Open Context has 1381 inbound links⁷ from sites outside its domain.⁸ However, one must be cautious in determining impact in this way, since different sites with different purposes will exhibit very different patterns in incoming and outgoing links. Open Context mainly publishes detailed inventories of artifacts and ecofacts that would be likely see loss without dissemination and access⁹. It has none of the more “glamorous” finds housed in major museums or key canonical texts central to a field of study¹⁰. Also, by publishing collections more than ontologies, Open Context will mainly be a consumer (rather than supplier) of outside Linked Data vocabularies, and will therefore have many outgoing links. As Linked Data approaches gain traction, sites like Open Context that mainly publish collections will see fewer inbound links than sites that publish Linked Data vocabularies.

4.4 Challenges

- *Building tools that require little technology expertise:* Over the course of this project, we demonstrated a variety of options for data dissemination, from sites targeted to specific communities (MicroCommons) to feeds of data aimed at enhancing existing sites (Badè Museum). All of these required some level of programming expertise, which was provided by this project’s Technology Team. Even “out of the box” tools such as those provided by Omeka seldom meet the full suite of needs expressed by producers and users of digital content and additional customization and programming help is usually needed to adapt a system to the specific requirements of a given community. There seems to be no way to get around requirements for some technical skills and expertise.
- *Working with student developers:* An area that we found at times extremely rewarding and at other times challenging was in working with student programmers. Perhaps even more than many of the senior participants on the project, students struggled to find time to dedicate to carrying out the technology developments. We attempted to minimize the amount of technology expertise needed to share heritage content, but even simple tools like Omeka (used for MicroCommons and BoneCommons) still require some degree of programming expertise, depending on how much the needs of the project diverge from the tools Omeka offers. More computer training for humanists (as well as humanities exposure for computer scientists) would help address this challenge.
- *Managing massive quantities of media:* The Kenan Tepe excavations recently submitted to Open Context a massive dataset documenting over 10 years of excavation. This dataset includes over 40,000 high resolution images (requiring more than 50 GB of storage). Managing large numbers of high resolution media files represents a challenge, in terms of server space, bandwidth, and metadata management. The Kenan Tepe team kept excellent metadata documentation of this massive image collection, and without such documentation, it would not be feasible for us to publish their excavation data in Open Context.

⁷ For Open Context, see: <http://siteexplorer.search.yahoo.com/search?p=opencontext.org&bwm=i&bwmo=d&bwmf=s>

⁸ The highly successful and significant Pleiades (<http://pleiades.stoa.org>) gazetteer, a central “hub” for linked data, has 1329 in-bound links. As Open Context and other projects reference Pleiades URIs more, Pleiades will see far more in-coming links. See: <http://siteexplorer.search.yahoo.com/search?p=pleiades.stoa.org&bwm=i&bwmo=d&bwmf=u>

⁹ Dryad, a site with a similar mission for sharing research data (in this case ecology) has about 460 inbound links. See: <http://siteexplorer.search.yahoo.com/search?p=datadryad.org%2F&bwm=i&bwmo=d&bwmf=u>

¹⁰ The Boston Museum of Fine Arts has 64,000+ inbound links. See: <http://siteexplorer.search.yahoo.com/search?p=www.mfa.org&bwm=i&bwmo=d&bwmf=u>. The wildly successful Perseus Digital Library has some 100,000+ inbound links. See: <http://siteexplorer.search.yahoo.com/search?p=www.perseus.tufts.edu%2F&bwm=i&bwmo=d&bwmf=u>

Nevertheless, this example illustrates the need for data dissemination services to prepare for increasingly large and complex submissions of data.

5. CONTINUATION OF THE PROJECT

As a result of the research undertaken by this project, Open Context has evolved to increasingly emphasize a model of “data sharing as publication.” Our outcomes show that professional researchers have data quality and “branding” concerns, as well as data documentation needs. Documenting datasets and improving data quality requires editorial supervision. Effective researcher data dissemination must involve some formal mechanisms of professional publication rather than informal Web 2.0 style sharing of user generated content. Furthermore, by establishing expectations and guidelines for data publication, we can promote documentation and sharing with the Web in mind—encouraging adherence to some simple practices that will facilitate discovery and repurposing of online content. Reflecting Open Context’s emphasis on formal data publication, in the summer of 2010, we established an [editorial board](#) comprised of domain experts representing several specializations in archaeology. In April 2011, we secured funding from the Alfred P. Sloan Foundation to integrate editorial supervision into data publication workflows for Open Context that will also have broader application in other disciplines.

“My dream tool is an open access web site that allows researchers to define a geographic area and search for all sites/collections by (1) chronological period and (2) material class, and then open links to reports, catalogs, maps, images and project participants. Catalogs are then able to be sorted and summarized.”

- Workshop Participant (Jan. 2009)

Asking researchers about their “dream tools” showed that many people want tools that will help them easily find high quality and relevant content. It takes time and effort to create high-quality data resources, and we believe that this will work will be distributed across the community, and across a variety of online publishing venues. As such, concerns about data portability and Web architecture become paramount. Thus, we have focused our project work to define and demonstrate Web architectures that are simple enough for wide adoption, yet capable of helping to meet researcher needs to find, aggregate, and analyze large-scale data resources drawn from multiple sources on the Web.

Finally, we are pursuing a number of related projects (such as the already-funded [Google Ancient Places](#) (GAP), with colleagues in the UK). Open Context works as a Linked Data provider to GAP, and the Mellon-funded Project Bamboo recently began work to incorporate methods pioneered by the GAP project into their services infrastructure. If successfully completed, Project Bamboo’s scaling of GAP will help generate larger and richer datasets that can be integrated with data from Open Context. In doing so, we hope to better integrate the archaeological record as represented in texts and publications with the archaeological record as represented in online data collections like Open Context.

International interest in Open Context has also increased and we have been invited on a series of collaborative proposals, including: using Open Context as a data publication model for Australian archaeology; using Open Context to share the digital contents of the German Archaeological Institute; and initiatives in Public Health (see below).

6. LONG-TERM IMPACTS

Our understanding of how to best deploy computing services in the humanities continues to evolve (Harley 2007; Harley et al. 2010). The guidelines developed over the course of this project will act as a useful starting point,

rather than a definitive last word, for user experience requirements in humanities computing. Other systems will have their own specific requirements. Nevertheless, this study will help jumpstart a more serious and systematic approach to the careful study, documentation, and dissemination of user experience requirements. As other initiatives similarly document and report on their own findings, the community will begin to understand essential requirements for successful computing platforms for humanities applications.

6.1 Direct Impacts Outside of the Humanities

We should also note that the technical developments of Open Context, a project primarily aimed to support humanistic data dissemination, have had a much wider impact than originally anticipated. Open Context offers a very generalized and abstracted data model, which is needed to support highly heterogeneous taxonomies and terminologies used in archaeology and related fields. This flexibility has opened opportunities to use Open Context's software in other disciplinary areas, including public health and environmental chemistry. This disciplinary breadth opens new funding opportunities, improving our sustainability prospects. Open Context's software now also powers the "Public Library of Materials" (Plum), a project sponsored by the UC Berkeley Center for Green Chemistry publishing information about chemical hazards (see: <http://plm.berkeley.edu/>).

7. GRANT PRODUCTS

Publications and presentations resulting from grant activities are listed in section 4.2 (Dissemination). The project activities were documented on the project's website, [Heritage Bytes](#). Significant developments to Open Context were made based on feedback gathered over the course of the project (listed in section 3.1.3) and these developments have led to newly funded directions, as described in Section 5. New or improved Web resources that resulted from this project are:

- Microcommons: <http://microcommons.org> (new resource, launched in June 2011)
- BoneCommons: <http://bonecommons.org> (improvements to functionality, based on research conducted as part of this grant)
- San Diego Archaeological Center: (new "datasets" page, with feed of SDAC content published in Open Context is now in beta testing at the San Diego Archaeological Center's own website).
- Badè Museum: <http://bade.psr.edu/content/tell-en-nasbeh-database> (new "Tell en-Nasbeh Database" page, with feed of Badè Museum content published in Open Context)
- Open Context: Since this project's inception, Open Context has published 9 additional archaeological datasets and has vastly expanded the range of content for the Domuztepe Excavations. We are also currently preparing another large scale excavation, Kenan Tepe (Directed by Bradley Parker, University of Utah), for publication in Open Context. Kenan Tepe's excavation documentation and media will approximately double the total collections in Open Context. The American Schools of Oriental Research also submitted data from ETANA-DL, a no-longer active NSF-funded project for republication in Open Context. Furthermore, in response to user needs for enhanced data quality and use of shared vocabularies, Open Context editors now use Google Refine to preprocess and clean data prior to publication in Open Context. Open Context now uses Linked Open Data methods to reference shared concepts, using URIs for concepts published by the Pleiades Gazetteer and the Encyclopedia of Life.

8. SUMMARY AND CONCLUSIONS

This project explored current technologies used by scholars working with cultural heritage data. The aim of the study was to determine how to better serve the needs of archaeologists working with primary data. We collaborated with creators and users of Web-based content and experimented with a variety of tools and information services. Our results show a wide diversity of user needs for sharing and presenting data to further

enhance organizational recognition, build community, and advance publishing goals. Users needed to selectively channel cultural heritage collections into a variety of contexts, especially book and journal settings, but also in less formal online conversational settings. Other users wanted to aggregate and analyze comprehensive datasets, gathered and centralized from a number of sources. Aggregation and reuse has strategic importance in justifying the whole data sharing enterprise. Sadly, this last need currently remains unmet because many relevant digital collections have made very poor use of Web technologies.

Critical user-needs cannot be addressed at the locus of individual websites or applications in isolation. We concluded that, rather than promoting one specific Web application or “front-end” technology or interface design, more attention should be given to back-end design of Web-based systems. Systems built following certain backend design principles (detailed by this project) will encourage greater longevity, exposure, and informed reuse. Rather than building isolated systems (“silos”), the scholarly community needs to better embrace the Web as rich source of contextualizing data and services. In doing so, access to “machine-readable” data will become the rule, rather than the exception. With enough professionally created machine-readable data, meaningful aggregation needed to sustain researcher interests can become feasible.

We recognize that sponsors of humanities collections have little interest in aligning their work toward good Web architectures merely for the sake of following abstract technical goals. Thus, we look toward future work to help demonstrate how Web architecture plays a critical role in enabling compelling humanistic research applications.

WORKS CITED

- Carr, Reg (2006) What Users Want: An Academic 'Hybrid' Library Perspective. Ariadne.
<<http://www.ariadne.ac.uk/issue46/carr/>> (Accessed January 12, 2008).
- Dempsey, Lorcan (2006) The (Digital) Library Environment: Ten Years After. Ariadne.
<<http://www.ariadne.ac.uk/issue46/dempsey/>> (Accessed March 4, 2008).
- Harley, Diane (2007) Why study users? An environmental scan of use and users of digital resources in humanities and social sciences undergraduate education. First Monday 11:2.
- Harley, Diane, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and C. Judson King (2010) *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Retrieved October 7, 2010 (<http://escholarship.org/uc/item/15x7385g>).
- Hearst, Marti. (2006) Clustering versus Faceted Categories for Information Exploration. *Communications of the ACM* 49(4):59-61. Retrieved February 25, 2009.
- Morville, Peter (2004) User Experience Design. Semantic Studios.
<<http://www.semanticstudios.com/publications/semantics/000029.php>> (Accessed March 10, 2008).
- Raben, Joseph (2007) Tenure, Promotion and Digital Publication. Digital Humanities Quarterly 1.
<<http://www.digitalhumanities.org/export/sites/default/vol/001/1/000006.html>> (Accessed March 1, 2008).
- Rubinoff, Robert (2004) How to Quantify the User Experience. Sitepoint.
<<http://www.sitepoint.com/print/quantify-user-experience>> (Accessed March 17, 2008).
- Warwick, Claire, Melissa Terras, Paul Huntington, and Nikoleta Pappa (2008) If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Lit Linguist Computing* 23:85-102.

APPENDICES

Appendix 1: The Workshop

Exploring User Needs Around Digital Heritage

UC Berkeley, School of Information, South Hall, Room 110

Agenda Highlights

Day 1: Friday, January 23, 2009

10 – 10:45: Interactive Issue Discussions: We'll use a novel and entirely active session format to address some of the framing issues of the meeting and invite each participant to weigh in with their unique perspectives and experiences.

10:45 – 11:15: Getting Acquainted: We'll work in both small-group and larger-group formats to explore several essential topics. Each participant will be encouraged to introduce themselves, and describe their communities of practice, as well as what kind of content they use and create.

11:30 – 12:45: Comparing Modes of Collaboration: We'll next move on to explore how each community is currently communicating and collaborating. In addition, the group will identify challenges, questions, and learnings about communication and collaboration.

1:45 – 3:00: "What's in your Toolbox?" This session will offer participants a chance to compare what technologies they're using in their work, both in terms of what works well and what falls short.

3:15 – 4:45: Break-Out Sessions

Primary Data, Instruction, and Public Audiences: Participants will share their thoughts on how access to primary content can enhance learning. What is needed to make primary data useful for instruction, what are the potential applications and challenges? How does improved access to content impact local and global communities and stakeholders?

Professional and Quality Concerns:

This session will discuss the professional and academic incentives around publishing datasets. What are the rewards and potential for sharing datasets? What are the risks? How may greater data sharing impact perception of quality? How can quality be assured and indicated?

Security, Privacy and Ownership: We will discuss the risks associated with greater transparency, in terms of looting and/or misuse of content. What safeguards should be put in place to balance openness with security concerns? What forms of recognition among different public communities are most desired, and how can they be achieved?

Day 2: Saturday, January 24, 2009

9:30 – 10:30: Break-Out Sessions:

Standards and Reporting Requirements: Some people work in contexts with strict requirements with regard to the use of different recording standards, data standards, etc. We will discuss various standards and requirements and how they are determined and enforced. We will also explore concerns about adhering to standards and how they might impact interpretive freedom and workflows.

Interoperability and Openness: There are many other sources of potentially relevant content important for research, instruction, and public engagement. What kinds of content and research materials from within and beyond the archaeological community would you like to use in relation to your own material? How would you use these other resources?

Stakeholders in Digital Heritage: This session will provide a forum for discussing how to engage with diverse communities related to our work. What are the main issues, needs, sensitivities, and claims that impact open dissemination?

11:00 – 12:15: Dealing with Diverse Content: Working in parallel, participants will enumerate the types of content that are important in their work, and prioritize them in terms of overall importance and significance. The discussion will span the spectrum of contents types, including multiple languages, gray literature, forums and images. Questions posed will include:

- * “What specifically do you do with each type of content?”
- * “How do you find what you’re looking for, both in terms of content you produce, and in terms of that which others produce?”
- * “In what ways might these processes be improved?”

1:15 – 2:30: Designing the Dream Tool: We’ll work collectively to posit an idealized solution for access and collaboration in each community. The goal will be not so much to specify a monolithic platform, but rather to tease out both shared and unique needs.

3:00 – 5:00: Where from Here? Each participant will identify an outcome for their community over the next two years. This may be identifying a research question that they think would be best addressed by collaborative research in their community. Or it may be a justification and potential next steps for their community to take to begin sharing content. Participants will brainstorm about potential collaborators from within their communities and ways to work together.

Workshop Participants

Coordinators:

Sarah W. Kansa, The Alexandria Archive Institute
Eric Kansa, UC Berkeley & Lead Developer of Open Context

Facilitator:

Allen Gunn, Aspiration

Documentation Team:

Noah Kersey, UC Berkeley
Sara van Wart, UC Berkeley
Francis Deblauwe, The Alexandria Archive Institute

Community Liaisons:

Stuart Campbell, University of Manchester
Catherine Foster, UC Berkeley & Badè Museum
Jim Allen, William Self Associates, Inc.
Eleanor Wilkinson, University of Durham
Dan Thompson, Global Heritage Fund
Levent Atici, University of Nevada, Las Vegas
Justin Lev-Tov, Statistical Research, Inc. & Zooarchaeologist
Carolina Mallol, Universidad de la Laguna, Tenerife, Spain
Margie Burton, San Diego Archaeological Center
Kelly Higelmire, San Diego Archaeological Center
Elizabeth Clevenger, Presidio Archaeology Lab (absent; interviewed at a later date)
Aaron Brody, Badè Museum (absent; interviewed at a later date)

Other participants:

Ahrash Bissell, CC Learn
Sophia Krzys Acord, UC Berkeley

Appendix 2: Small Working Groups

Group I: Data-Heavy Institutional Collections

Participants: Kelly Higelmire, Margie Burton, Liz Clevenger, Aaron Brody and Catherine Foster

Overview and Outcomes: This working group explores how open technologies can best meet the diverse needs of researchers dealing with data-heavy archaeological content from an institutional perspective. We are working with three groups, two from California archaeology (the San Diego Archaeological Center and the Presidio Trust in San Francisco) and one based in California but with institutional collections from the Middle East (the Badè Museum). While all organizations seek to find ways to leverage digital technologies to better organize and share their content, their choices are influenced by different aims and challenges. By achieving a better understanding of their various needs, we aim to create tools for content-sharing that are useful to all groups and discuss solutions to maximize the reach, discoverability and preservation of heritage content. One great challenge in the institutional context is that, unlike the individual researcher seeking to publish her data somewhere on the Web, institutions generally have a Web presence and want to integrate their Web-published collections seamlessly into their Web sites.

The three institutions are as follows:

San Diego Archaeological Center:

Content: Archaeological objects (descriptions, images and supporting documentation) from several sites in the San Diego region (potentially great breadth across a large area and multiple time periods)

Needs: To improve exposure and use of their collections on their own website to encourage student and private research

Main concern: Developing an online presence while maintaining site security and cultural sensitivity.

Presidio Archaeology Lab:

Content: Images, maps, artifacts, analyses and reports from one, mainly historic period, excavation area in the Presidio of San Francisco (Building 39).

Needs: Want to grow their “lay” audience through digital platforms and want to see comprehensive and faster access to content for research and educational purposes and to improve public outreach.

Main concern: Ensuring that they have some level of control over the shared content so that the quality of the content is maintained

Badè Museum:

Content: Images and catalog cards for artifacts from multiple time periods at Tall en-Nasbeh in Palestine.

Needs: Want to provide better access to this collection so that researchers from elsewhere can explore its contents and use it for research.

Main concern: Want the collection available through their institutional website to increase their public exposure.

Planned Presentations / Publications:

- Exploring User Needs in Archaeological Data-Sharing: Case Studies from California Archaeology (conference presentation by Sarah Whitcher Kansa, Kelly Higelmire, Liz N. Clevenger and Eric C. Kansa)
- Googling the Grey: Open Data, Web Services, and Semantics (by Eric C. Kansa, Sarah Whitcher Kansa, Margie M. Burton, Cindy Stankowski) Published in 2010 in [*Archaeologies, Journal of the World Archaeological Congress*](#)

- Tell en-Nasbeh's Digital Rebirth (by Aaron Brody, Catherine Foster and Eric Kansa)

Group II: Blind Data Analysis and Report Writing Study

Participants: Justin Lev-Tov, Levent Atici and Sarah Whitcher Kansa

Overview and Outcomes: The Web has great potential for supporting low-cost dissemination of large and complex researcher datasets; however, most scholars perceive few rewards for sharing their primary data. Given that documenting a dataset for reuse requires effort, researchers feel little incentive to share their data. To help explore where documentation effort can be most effectively spent, this paper considers data sharing from the perspective of "end users" who consume legacy data for reuse and analysis. This group undertook a blind analysis and report-writing study, using the publicly available faunal dataset from excavations at Chogha Mish, Iran, in the 1960s and 1970s. The group followed a few common ground rules, but individually defined and customized analytical methods and interpretive aims. Despite widely diverse interpretations, the analysts came to similar conclusions about the dataset's utility to explore general topics. This blind analysis shows how data "integrity" (suitability for analysis) plays a central and under-recognized role in a dataset's potential for reuse. Though sparsely documented, the dataset used in this study showed sufficient integrity (judged by the three analysts) to merit further analysis. Supplementary documentation was achieved using the published literature as well as inferences based on the dataset itself. This study concludes with recommended best publishing practices to promote the sharing of analytically useful datasets.

Planned Presentations and Publications:

- "Other People's Data: Blind Analysis and Report Writing as a Demonstration of the Imperative of Data Publication"
- "New Light on Diet and Animal Use at Choga Mish, Iran: Analysis on Faunal Data from the 1972-1978 Excavations"

Group III: Media-Heavy, Researcher-Managed Collections Working Group

Participants: Eleanor Wilkinson, Catherine Foster, Carolina Mallol and Sarah Whitcher Kansa (with input from Nada Shabout, Saleem Al-Bahloly, and Heather Ford)

Overview and Outcomes: An increasing number of institutions and large projects are sharing their collections online (as demonstrated by the continuing series of project reviews on this blog). These entities tend to find enough (but often not much) funding, time and technical expertise to share at least a portion of their digital content. Large institutions like museums and well funded research groups often develop their own customized systems for publishing collections online. Such customized approaches are typically out of reach for smaller organizations and individuals.

With this concern in mind, part of our user experience study has been to target individuals and small organizations to learn about their content-sharing needs and to develop tools that they can easily adopt to better expose their collections. In our user experience workshop in January 2009, four researchers identified bodies of content they seek to share or digital tools they desire to improve communication and collaboration among their archaeological sub-disciplines. From this, we formed the "Researcher-Managed Collections Working Group," aimed at determining the content-sharing, access and management needs of individual researchers who seek to

share digital resources that they have collected about a specific place, region or material type. A key goal was to attempt the collaborative development of online reference collections.

The four projects involve zooarchaeology, microdebris, micromorphology and content from one region of northern Iraq. These resources tend to be image-heavy and the communities they serve seek both to contribute their own content and enrich the existing content through commentary. More or less, many have the goal of publishing reference collections, and this goal seems to emphasize media and discussion more than the results of data analysis (as emphasized by Open Context). They also desire simple tools for having discussions and, by involving a global community, moving toward standards or consensus in their discipline. Above all, the creators of these resources have limited funding and technical support, so the system needs to be simple and ideally, free.

Four projects are underway, and all are experiment with a combination of Omeka and Open Context. Features of this hybrid system include: multiple language capabilities, faceted browsing to expose the system's content, citation, licensing for maximum reuse, and community input through tagging and image annotation. However, the participants have focused on different aspects of the two systems, depending on the needs of their different communities. In the upcoming working group meeting, participants will meet to discuss their progress, see how others have visualized content using the same tools, and define steps toward having a final product to release to their interested communities.

Planned Presentations / Publications / Websites:

- Beyond BoneCommons: Recent Developments in Zooarchaeological Data Sharing (conference presentation at the International Council for Archaeozoology, and published in January 2011 in *The SAA Archaeological Record*, along with eight other papers related to digital tools for improving research in archaeology (special issue edited by Sarah Witcher Kansa and Iain McKechnie; available online at: <http://www.alexandriaarchive.org/bonecommons/items/show/1731>)
- "An open source content dissemination system" (presented by Carolina Mallol to the micromorphology community, June 2009)
- The Modern Art Iraq Archive: Web tools for Documenting, Sharing and Enriching Iraqi Artistic Expressions (presented by Sarah Witcher Kansa at the Digital Humanities 2010 conference) [Note: This presentation was funded by a Digital Humanities Start Up Grant (HD-50821) but is relevant here because it uses Omeka tools. The MAIA site can be accessed here: <http://artiraq.org>]
- Microcommons: <http://microcommons.org>
- BoneCommons: <http://bonecommons.org>

Group IV: Large Excavation Datasets Working Group

Participants: Stuart Campbell, Levent Atici and Benjamin Porter

Overview and Outcomes: This working group addresses the challenges around sharing large amounts of data from multi-year research projects involving many participants from a variety of institutions. For existing projects, can we streamline the data publication process to minimize the amount of work project participants have to do to share their content? Can we provide advice for existing projects on how to "clean up" their content so it can be more easily shared and understood? For new projects, can we establish some guidelines to help them plan for simple future dissemination of project content?

Planned outcomes of this project include a set of guidelines to help data authors to prepare their content for Web publication, and technological developments to Open Context to improve data ingest, documentation, discovery,

and citation. We also planned and completed publication of several new datasets in Open Context, together with enhancements to workflows and software tools to improve data documentation and initial steps toward implementing Linked Data methods.

We also explored incentive concerns with regard to data sharing. Although these are well documented in related studies (see Harley et al. 2010:14, 30-32), our self-selecting group of project participants had fewer reservations about sharing data. They saw themselves as somewhat “cutting edge” in their interest in data sharing, and wanted data sharing efforts to augment and enhance their other publishing goals. We experimented with some features like widgets (<http://opencontext.org/widget>) to dynamically feed information about data publication and use (counts of views) to researcher’s online CVs and other websites, but these proved to be less of interest than clear forms of bibliographic citation for Open Context datasets. We are currently working with the California Digital Library to mint stable archival identifiers and implement the DataCite standard for such citations.

Planned Publications:

We explored models for incorporating published datasets with narrative publication. We tested these models in collaboration with Sorin Matei (Purdue University) and his publishing platform, *VisiblePast.net*. We are also currently processing data from the Kenan Tepe excavations (directed by Bradley Parker, University of Utah). To help document and disseminate our methods and perspectives for publishing large, multi-specialist archaeological datasets, we have completed several publications in peer-review journals, online only venues, and university press-published edited volumes (see section 4.2).

Appendix 3: Website Reviews

This project's Dissemination Coordinator, Francis Deblauwe, reviewed 68 websites between February 2009 and December 2010. Of these, 25 were sites sharing archives, databases, or other digital content related to ancient cultural heritage. The remaining 43 sites were all related to the cultural heritage of Iraq (for use in a case study on developments and trends in Iraq-related digital content). For each website, Deblauwe addressed the review criteria below.

Appendix 3.1: Review Criteria

- I. Site name
URL
Main content offered
Authorship
Where is it housed? Who maintains it?
Permanence and Archiving
Licensing: Is there clear licensing indicated? Is it for each item or for the site in general? Is it clear how you can use (and cite) items in this resources?
- II. Address each of the following user experience criteria for the site (after Morville 2004:Fig. 2):
 1. Usefulness: What needs does the system meet? Is it filling a niche? Does it house substantive content? Discuss the depth and breadth of the content in the system and how useful it appears to be for its intended audience.
 2. Ease of Use: Is the system easy to use? How do you navigate around the site? Is it easy to find what you're looking for?
 3. Appealing: Is the system well designed? Is it professional-looking?
 4. Accessibility: Is the system itself easy to find? Are items from it easy to find?
 5. Credibility: Is the site professional? Does it hold good research content? Most importantly, does it provide clear and useful options for citation and licensing?
 6. Reuse: Is it easy to export data in convenient formats? That is can data be processed, reused? (e.g. Excel, comma-delimited text, XML—NOT pdf, and not just Web pages)

Appendix 3.2: Reviewed Sites

Sites Reviewed for the General Landscape Survey:

Project Name	Project URL
Animal Bone Metrical Archive Project (ABMAP)	http://ads.ahds.ac.uk/catalogue/resources.html?abmap
Archaeology Data Service (ADS)	http://ads.ahds.ac.uk/
Ancient Mesopotamia: This History, Our History	http://mesopotamia.lib.uchicago.edu
Ashes2Art	http://www.coastal.edu/ashes2art/
American Theological Library Association(ATLA) Cooperative	http://www.atla.com/cdri/

Digital Resources Initiative	
Big Digs Go Digital	http://www.bigdigs.org
British School at Rome (BSR) Library and Archive Digital Collections	http://www.bsrdigitalcollections.it/
Çatalhöyük Databases On-Line	http://www.catalhoyuk.com/database/catal/
Cuneiform Digital Library Initiative	http://cdli.ucla.edu/
CyArk	http://www.cyark.org
Digital Archaeological Archive of Comparative Slavery	http://www.daacs.org
Digital Archaeological Atlas of the Holy Land	http://daahl.ucsd.edu/
Digital Archive Network for Anthropology and World Heritage (DANA-WH)	http://www.dana-wh.net/
Digital Corinth SD	https://k20.internet2.edu/projects/25
Electronic Tools and Ancient Near East Archives (ETANA)	http://www.etana.org
Foundation for the Advancement of Mesoamerican Studies, Inc. (FAMSI)	http://www.famsi.org
Hermitage APP	http://www.hermitagemuseum.org/fcgi-bin/db2www/browse.mac/category?selLang=English
The Louis Shotridge Digital Archive	http://penn.museum/collections/shotridge/about_the_project.html
National Archaeological Database (NADB)	http://www.nps.gov/history/archeology/TOOLS/Nadb.htm
Nautical Archaeology Digital Library	http://nadi.tamu.edu
Open Context	http://www.opencontext.org/
Open Research Scan Archive (ORSA)	http://plum.museum.upenn.edu/~orsa/
PaleoDB	http://paleodb.org/cgi-bin/bridge.pl
Perseus 4.0	http://www.perseus.tufts.edu/hopper/
The Digital Archaeological Record (tDAR)	http://www.tdar.org

Sites Reviewed for the Iraq Case Study:

Project Name	Project URL
Open Modern Art Collection of Iraq (OMACI)	http://omeka.ischool.berkeley.edu/artiraq/
The Looting of the Iraq National Museum	http://www.theartnewspaper.com/iraqmus/
Le Musée d'Irak - Iraq Museum	http://www.ezida.com/musee0.htm
Ashurbanipal Library Project	http://www.fincke.uni-hd.de/nineveh/index.htm
Digital Map of the Ancient Near East	http://Web.archive.org/Web/20070608230013/http://www.cba-inst.org/Pages/DigiMap.html%20
Bureau for Recovering and Investigating Iraqi Looted Antiquities (BRILA)	http://www.centroscavitorino.it/en/progetti/iraq/brila.asp
Opere d'arte rubate (Iraq section)	http://www.carabinieri.it/carabinieri/?MIval=tpc_presite&MIqry2pass=on&iraq=on

The Virtual Museum of Iraq	http://www.virtualmuseumiraq.cnr.it ; http://seattletimes.nwsources.com/html/travel/2008785488_Webiraqmuseum25.html ; http://www.neh.gov/DigitalHumanities/Conference_07Oct/Cultraro.pdf
Electronic Cultural Atlas Initiative (ECAI) Iraq	http://ecai.org/iraq/ ; http://www.dlib.org/dlib/may04/05featured-collection.html ; http://articles.latimes.com/2003/jun/02/local/me-museum2
Babylonian Texts of the First Millennium B.C.	http://Web.t-online.hu/enkidu/ ; http://enkidu.iWeb.hu/
The Iraq Heritage Program - Saving Our Cradles of Civilization	http://opencontext.org/projects/GHF1PRJ0000000025
Global Heritage Network Beta (Iraq sites)	http://ghn.globalheritagefund.org/
Google	http://www.nytimes.com/2009/11/25/world/middleeast/25iraq.html?_r=1&ref=technology ; http://www.theiraqmuseum.com
Knowledge and Power in the Neo-Assyrian Empire	http://kn.prs.heacademy.ac.uk/
Emergency Red List of Iraqi Antiquities at Risk	http://icom.museum/redlist/irak/en/index.html
International Network for Contemporary Iraqi Artists (iNCiA)	http://www.incia.co.uk/
Endangered Iraqi Objects	http://www.interpol.int/Public/WorkOfArt/Iraq/objetsEnDanger/Default.asp
Stolen Iraqi items	http://www.interpol.int/Public/WorkOfArt/Iraq/Items.asp
Iraq Museum Int'l (formerly Baghdad Museum Project)	http://www.baghdadmuseum.org/
Digital Hammurabi	http://www.jhu.edu/digitalhammurabi/
Kurdish Textile Museum	http://www.kurdishtextilemuseum.com/
The Northwest Palace of Ashur-nasir-pal II at Nimrud	http://www.learningsites.com/NWPalace/NWPalhome.html
Minerva Exclusive: National Museum of Iraq	http://Web.archive.org/Web/20070609023603/http://minervamagazine.com/exclusives/iraq_01.php
ImageServer, Iraq section [1991 Gulf War stolen artifacts]	Available via Internet Archive's Wayback Machine here
The Neo-Babylonian Trial Procedure	http://www.sju.edu/~bwells/nbtp/index.htm
The Iraq Museum Website	http://www.theiraqmuseum.org/
Archive of Mesopotamian Archaeological Reports (AMAR)	http://libmedia.cc.stonybrook.edu/amar/
Duplicazione e Rinascita	http://www.uplink.it/assiriologiaprogetto/ ; http://www.archaeogate.org/vicino_oriente/article/320/1/progetto-dr-duplicazione-e-rinascita-progetto-di-catalo.html http://www.uplink.it/assiriologia/Progetti.asp?idp=3
Iraq-Museum Baghdad	http://Web.archive.org/Web/20050324020423/ http://info.uibk.ac.at/c/c6/c616/museum/museum.html
Iraq War & Archaeology (IW&A)	http://iwa.univie.ac.at/
Assyrian Empire Builders	http://www.ucl.ac.uk/sargon/
Cuneiform Digital Palaeography Project (CDP)	http://www.cuneiform.net/
Digital Nineveh Archives	http://www.digitalnineveharchives.org/ ; http://archive.cyark.org/nineveh-region-intro
Digital Corpus of Cuneiform Lexical Texts (DCCLT)	http://cdl.museum.upenn.edu/dcclt/index.html ; http://www.neh.gov/news/archive/20040408.html ; http://www.ocf.berkeley.edu/~veldhuis/research.html ;

Cuneiform Digital Library Initiative (CDLI)	http://cdli.ucla.edu/ ; http://www.neh.gov/news/archive/20040408.html ; http://www.neh.gov/news/humanities/2002-09/gilgamesh.html
The Threat to World Heritage in Iraq	http://users.ox.ac.uk/~wolf0126/
Diyala Project (Diyala Virtual Archive)	http://oi.uchicago.edu/research/projects/diy/ ; http://oi.uchicago.edu/gallery/diyala/ ; http://chronicle.uchicago.edu/041202/diyala.shtml
Lost Treasures from Iraq (Iraq Museum Database)	http://oi.uchicago.edu/OI/IRAQ/iraq.html
Electronic Text Corpus of Sumerian Literature (ETCSL)	http://www-etcsl.orient.ox.ac.uk/ ; http://www.csad.ox.ac.uk/CSAD/Newsletters/Newsletter7/Newsletter7b.html
Electronic Pennsylvania Sumerian Dictionary Project (ePSD)	http://psd.museum.upenn.edu/
State Archives of Assyria Online (SAAo)	http://cdl.museum.upenn.edu/saa/index.html ; http://www.helsinki.fi/science/saa/cna.html ; http://cdl.museum.upenn.edu/cgi-bin/cdlpager?project=saa
Middle Eastern Geodatabase for Antiquities (MEGA)-Iraq	http://getty.museum/conservation/field_projects/iraq/index.html ; http://www.neh.gov/news/archive/20040408.html ; http://www.neh.gov/news/humanities/2004-11/lostempire.html
Iraq ReCollection	http://www.library.yale.edu/digIraq/

Appendix 3.3: Archaeological Collections Offering Machine-Readable Data

- *Project Bamboo*: A major digital humanities project, Project Bamboo published documentation stating: "...this will enable integration with applications and environments that can 'speak' RESTful web service protocols or consume and serve Atom feeds."¹¹ Bamboo's adoption of Atom will help speed adoption.
- *Portable Antiquities Scheme*: The Portable Antiquities Scheme (PAS) is a government-sponsored voluntary scheme to record archaeological objects found by members of the public in England and Wales. It currently records over 240,000 objects, plus an additional 78,600 Roman era coins, each of which is individually addressable by a stable URI. Its searches can be returned as Atom feeds. Many PAS objects, while found in the United Kingdom, were produced in the Mediterranean region. Among the relevant origins that can be easily searched are: Alexandria, Rome, Syria, Nicomedia (Izmit, Turkey), and Heraclea (Eregil, Turkey). Elements of the PAS database are already accessible as Atom-feeds and we expect this capability will expand.
- *Museum Databases*: American and European museums increasingly publish their collections databases online. Many of these already provide URIs for individual objects. These collections would better support research if they offered services for machine interactions and aggregation.
- *Arachne*: Arachne is a major online collection containing some 250,000 objects published by the University of Cologne and German Archaeological Institute. Arachne has long offered machine-readable data and aligned its collection with community semantic standards (including the CIDOC-CRM). Developed in consultation with project participant Eric Kansa, Arachne now offer query defined Atom-feeds..
- *Field Project Databases*: The Athens Agora project has published large quantities of primary archaeological data online, also making these available in JSON and XML.
- *tDAR*: Project participant Eric Kansa is collaborating with Adam Brin, tDAR's lead developer to share "Open Search" services (based on Atom feeds) between Open Context and tDAR, the archaeological data developed and maintained by Digital Antiquity.
- *Nomisma.org*: Nomisma.org is a site developed by Sebastian Heath (NYU / American Numismatic Society) for documenting coin hoards in the ancient Mediterranean world. Nomisma offers machine-readable data

¹¹ See: <http://www.projectbamboo.org/infrastructure/> (accessed May 14, 2011).

in XHTML+RDFa and RDF formats. This site represents an exemplar implementation of “Linked Open Data” for archaeology and Classics.

- *MANTIS*: The American Numismatic Society recently launched MANTIS and online collections database of ancient and modern coins from across the globe. MANTIS offers machine-readable representations of searches (in Atom) and of individual coins in XML.
- *Pleiades*: Pleiades offers stable URIs to place, location, and toponymn entities, helping to make it a key reference source and “Linked Open Data” vocabulary. The Pleiades Gazetteer also represents a key reference implementation of RESTful Web services, expressing its content in machine-readable JSON, Atom, and KML formats.